RESEARCH ARTICLE

AI-Assisted Optimization of Polyherbal Formulations for Diabetes Mellitus Therapy

Veerabathiran Balasubramanian[1], Neelambari S[2], Komal Verma[3], Rajesh KS[4], Shikha Saxena[5], Ramanpreet Walia[5], Shoheb Shakil Shaikh[6] and Swati Madan[5]*

[1]JOY University, School of Pharmacy, Tirunelveli, Tamil Nadu – 627116, India.

[2]Dr. M.G.R. Educational and Research Institute, Velappanchavadi, Chennai – 600077, India.

[3]School of Pharmacy, Babu Banarasi Das University, 111, Faizabad Rd, Atif Vihar, Lucknow, Uttar Pradesh – 226028, India.

[4]Nitte (Deemed to be University), NGSM Institute of Pharmaceutical Sciences (NGSMIPS), Department of Pharmacology, Mangaluru – 575018, India.

[5]Amity Institute of Pharmacy, Amity University, Amity Rd, Sector 125, Uttar Pradesh – 201301, Noida, India.

[6]RSM'S NN Sattha College of Pharmacy, Ahilyanagar, Maharashtra, India – 414001.

***Corresponding Author:**

Dr. Swati Madan,

Professor,

Amity Institute of Pharmacy, Amity University,

Amity Rd, Sector 125, Uttar Pradesh – 201301, Noida, India

Email: smadan3@amity.edu

Abstract

Diabetes mellitus (DM) represents a major global non-communicable disease characterized by persistent hyperglycemia and multifactorial metabolic dysfunction. Although synthetic antidiabetic drugs remain clinically effective, long-term therapy is often associated with adverse effects, limited target specificity, and poor patient adherence. Polyherbal formulations derived from medicinal plants offer synergistic and multitargeted therapeutic potential; however, optimization of such complex combinations remains scientifically challenging. The present study employed artificial intelligence (AI)-assisted computational approaches, including machine learning, network pharmacology, molecular docking, and ADMET prediction, to identify and optimize polyherbal combinations for diabetes management. Eight ethnopharmacologically validated medicinal plants were selected, and phytochemical data were retrieved from PubChem, ChEMBL, IMPPAT, and TCMSP databases. Random forest and artificial neural network models were trained using curated bioactivity datasets to predict antidiabetic efficacy. Molecular docking against key diabetic targets, including DPP-4, PPARγ, α-glucosidase, AMPK, and IRS-1, demonstrated strong binding affinities ranging from −7.4 to −9.8 kcal/mol. Among 347 screened phytoconstituents, 68 satisfied drug-likeness criteria, while the optimized combination of berberine, gymnemic acid, charantin, and curcumin exhibited superior predicted therapeutic potential. Network pharmacology identified multiple enriched pathways associated with insulin signaling and glucose homeostasis. ADMET analysis confirmed favorable pharmacokinetic and toxicity profiles. Overall, this study establishes an evidence-based, non-animal computational framework for precision polyherbal formulation development in diabetes therapy.

Keywords: Artificial intelligence; polyherbal formulations; diabetes mellitus; machine learning; network pharmacology; molecular docking; phytoconstituents; precision herbal medicine.

How To Cite This Article: Balasubramanian V, Neelambari S, Verma K, Rajesh KS, Saxena S, Walia R, Shaikh SS, Madan S. AI-Assisted Optimization of Polyherbal Formulations for Diabetes Mellitus Therapy. Int J Drug Deliv Technol. 2026;16(49s): 1074-1091. DOI: 10.25258/ijddt.16.49s.124

1. INTRODUCTION

1.1 Global Burden of Diabetes Mellitus

Diabetes mellitus constitutes a metabolic syndrome of heterogeneous etiology characterized by chronic hyperglycemia resulting from absolute or relative insulin deficiency, impaired insulin secretion, or both. According to the International Diabetes Federation (IDF), approximately 537 million adults worldwide were living with diabetes in 2021, with projections indicating a catastrophic escalation to 783 million by 2045.[1] The disease exacts an enormous socioeconomic toll, accounting for over USD 966 billion in global healthcare expenditure annually.[2]

Particularly in low- and middle-income countries, the epidemiological transition toward sedentary lifestyles, energy-dense dietary patterns, and rapid urbanization has accelerated the incidence of type 2 diabetes mellitus (T2DM), which constitutes approximately 90–95% of all diabetes cases.[3]

The pathophysiology of T2DM involves a complex interplay of insulin resistance, progressive pancreatic beta-cell dysfunction, dysregulated incretin signaling, and systemic low-grade inflammation.[4] Chronic hyperglycemia precipitates advanced glycation end-product (AGE) accumulation, oxidative stress, and mitochondrial dysfunction, culminating in microvascular and macrovascular complications including nephropathy, retinopathy, neuropathy, and cardiovascular disease.[5] These complications substantially diminish quality of life and place sustained pressure on primary and secondary healthcare infrastructures globally.

1.2 Limitations of Conventional Antidiabetic Therapies

The contemporary pharmacotherapy of T2DM encompasses diverse drug classes including biguanides (metformin), sulfonylureas, thiazolidinediones, sodium-glucose cotransporter-2 (SGLT-2) inhibitors, dipeptidyl peptidase-4 (DPP-4) inhibitors, and glucagon-like peptide-1 (GLP-1) receptor agonists.[6] While these agents demonstrate measurable glycemic efficacy, their long-term use is associated with a constellation of adverse effects. Metformin, though considered first-line therapy, is contraindicated in chronic kidney disease and associated with gastrointestinal intolerance in up to 30% of patients.[7] Sulfonylureas carry risks of hypoglycemia and weight gain, while thiazolidinediones have been linked to fluid retention and cardiovascular complications.[8]

Furthermore, the progressive and multifactorial nature of T2DM pathogenesis demands combination pharmacotherapy that addresses multiple dysfunctional axes simultaneously. Monotherapy invariably yields suboptimal glycemic control over time due to disease progression and beta-cell exhaustion. Therapeutic inertia, polypharmacy burdens, economic inaccessibility particularly in developing nations, and the persistent unmet need for agents targeting the underlying inflammatory and oxidative pathophysiology underscore the imperative for novel therapeutic strategies.

1.3 Polyherbal Therapy in Diabetes: Rationale and Synergy

Herbal medicine has sustained an uninterrupted therapeutic legacy across diverse cultural traditions for millennia, and numerous phytoconstituents have demonstrated compelling antidiabetic properties in mechanistic investigations. Medicinal plants exert antidiabetic effects through multiple complementary mechanisms including alpha-glucosidase and alpha-amylase inhibition, potentiation of insulin secretion, improvement of peripheral insulin sensitivity, protection of pancreatic beta-cells from oxidative damage, attenuation of hepatic glucose output, and modulation of adipokine signaling.[9]

Polyherbal formulations (PHFs)—preparations combining two or more medicinal plants—capitalize on the principle of pharmacological synergism, wherein the combined therapeutic effect exceeds the arithmetic sum of individual plant contributions. Such synergism arises from phytochemical interactions at shared molecular targets, complementary modulation of distinct pathophysiological axes, and pharmacokinetic enhancement including improved bioavailability and metabolic stability through plant matrix effects.[10] For instance, the combination of Gymnema sylvestre with Momordica charantia has been reported to exhibit additive effects on glucokinase activation and glucose transporter (GLUT-4) expression enhancement.[11] The concept of 'network pharmacology,' wherein multiple phytoconstituents simultaneously modulate interconnected disease-associated protein networks, provides a sound mechanistic rationale for PHF-based interventions.

1.4 Challenges in Polyherbal Formulation Optimization

Despite their therapeutic promise, the rational development of optimized PHFs confronts substantial methodological challenges. The chemical complexity of polyherbal systems—encompassing potentially hundreds of bioactive compounds across constituent plants—renders systematic evaluation of all possible combinations experimentally intractable using conventional approaches.[12] Variability in phytochemical profiles due to geographical, seasonal, and post-harvest factors introduces reproducibility challenges. The risk of pharmacokinetic and pharmacodynamic herb-drug or herb-herb interactions demands rigorous evaluation.[13]

Traditional pharmaceutical development paradigms rely on reductionist single-compound, single-target models that fail to capture the emergent properties of complex herbal mixtures. Quality standardization, determination of optimum synergistic ratios, identification of lead phytoconstituent combinations, and prediction of their ADMET profiles collectively represent critical bottlenecks in translating ethnopharmacological knowledge into clinically validated PHFs.[14] These limitations call for advanced computational methodologies capable of

processing high-dimensional chemical and biological data to rationally guide formulation design.

1.5 Artificial Intelligence in Pharmaceutical Sciences

The convergence of artificial intelligence with pharmaceutical sciences has inaugurated a paradigm shift in drug discovery and development. AI encompasses a broad spectrum of computational methods including machine learning (ML), deep learning (DL), natural language processing (NLP), and reinforcement learning, all of which have found impactful applications across the drug discovery pipeline.[15] In preclinical drug discovery, ML algorithms trained on large-scale bioactivity datasets have demonstrated remarkable predictive accuracy for target engagement, off-target effects, and pharmacokinetic parameters.[16]

Random forest (RF) classifiers, support vector machines (SVM), gradient boosting algorithms, and graph neural networks (GNNs) have been successfully applied to virtual screening, lead optimization, and QSAR modeling.[17] Deep learning architectures including convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have demonstrated superior performance in molecular property prediction and de novo molecular generation tasks.[18] The integration of these AI modalities with quantum chemical descriptors, topological fingerprints, and structural similarity metrics enables multidimensional characterization of phytochemical libraries with unprecedented computational efficiency.

1.6 Machine Learning and Computational Pharmacology

Computational pharmacology represents the application of mathematical, statistical, and computational tools to model drug-target interactions, predict pharmacological outcomes, and optimize molecular structures.[19] QSAR (Quantitative Structure-Activity Relationship) modeling, which relates molecular structural descriptors to biological activity, constitutes a foundational pillar of computational drug design. Modern QSAR approaches integrate diverse molecular fingerprints, Morgan fingerprints, MACCS keys, and ECFP descriptors with ensemble learning algorithms to achieve robust activity predictions across chemical spaces.[20]

Network pharmacology, an emerging systems biology-based approach, maps multi-component, multi-target interactions within disease-associated molecular networks to decode the holistic mechanisms of complex therapeutic agents including herbal medicines.[21] This approach is particularly well-suited to polyherbal systems, as it models the disease interactome as an interconnected protein-protein interaction (PPI) network and identifies critical nodes through which phytoconstituents exert therapeutic modulation. The integration of network pharmacology with molecular docking and ML-based activity prediction provides a powerful multi-layered computational framework for PHF optimization.

1.7 AI in Herbal Drug Discovery

The application of AI specifically to herbal drug discovery has witnessed remarkable growth since 2018. Natural language processing algorithms have been deployed to mine traditional medicine literature and extract phytochemical-activity associations from unstructured bibliographic databases.[22] ML models trained on phytochemical datasets have successfully predicted anti-inflammatory, antidiabetic, and antioxidant bioactivities with high specificity.[23] AI-driven dereplication strategies have accelerated the identification of known bioactive scaffolds within complex plant extracts, enabling rational prioritization for downstream pharmacological investigation.

Specifically in diabetes research, AI-assisted virtual screening of phytochemical libraries against DPP-4, alpha-glucosidase, AMPK, PPARγ, and other validated antidiabetic targets has yielded promising lead compounds with favorable predicted potency profiles.[24] Furthermore, generative adversarial network (GAN)-based molecular generation models have been applied to design novel phytochemical derivatives with enhanced target selectivity and metabolic stability.[25] These developments collectively validate the transformative potential of AI in accelerating phytopharmaceutical development.

1.8 Research Gap and Novelty

Despite the accumulation of individual computational studies examining specific phytoconstituents or single herbal species against diabetes targets, a comprehensive AI-integrated framework for the systematic optimization of multicomponent polyherbal combinations accounting for synergistic interactions, network pharmacology, and comprehensive ADMET prediction simultaneously remains underexplored in the literature.[26] Most existing studies either evaluate single compounds in isolation or employ conventional combinatorial approaches without leveraging the full predictive power of modern ML architectures.

The present study addresses this gap by developing an integrated computational pipeline that combines phytochemical database mining, ML-based bioactivity prediction, network pharmacology analysis, molecular docking, and ADMET profiling to rationally identify and optimize a synergistic polyherbal combination for T2DM

therapy. This constitutes a methodologically novel, non-animal, and ethically rigorous approach to PHF development.

1.9 Study Objectives and Hypothesis

The primary objective of this investigation is to develop and validate an AI-assisted computational framework for the identification, evaluation, and optimization of polyherbal combinations targeting pathophysiological mechanisms of T2DM. The specific objectives are: (i) to construct and curate a comprehensive phytochemical database from eight antidiabetic medicinal plants; (ii) to apply ML classifiers to predict the antidiabetic bioactivity of retrieved phytoconstituents; (iii) to perform network pharmacology analysis to identify shared therapeutic targets and enriched biological pathways; (iv) to execute molecular docking studies against validated antidiabetic protein targets; (v) to conduct ADMET profiling to assess drug-likeness and pharmacokinetic suitability; and (vi) to identify the optimal polyherbal combination based on integrated AI-driven scoring. The central hypothesis posits that AI-assisted multi-parametric optimization will identify synergistic polyherbal combinations with superior predicted antidiabetic efficacy, favorable ADMET characteristics, and multimodal mechanisms compared to individual plant constituents.

2. MATERIALS

2.1 Selection of Antidiabetic Medicinal Plants

Eight medicinal plants with well-established ethnopharmacological use in diabetes management and documented scientific evidence of antidiabetic activity were selected for this investigation. The selection criteria encompassed: (i) documented antidiabetic efficacy in peer-reviewed scientific literature; (ii) availability of characterized phytochemical profiles in public databases; (iii) ethnopharmacological use in traditional medicine systems including Ayurveda, Siddha, Unani, and Traditional Chinese Medicine; and (iv) availability of pharmacological target data relevant to T2DM pathophysiology.

The eight plants selected were: Gymnema sylvestre (R.Br.) Schult. (Asclepiadaceae), Momordica charantia L. (Cucurbitaceae), Trigonella foenum-graecum L. (Fabaceae), Tinospora cordifolia (Willd.) Miers (Menispermaceae), Berberis aristata DC. (Berberidaceae), Ocimum sanctum L. (Lamiaceae), Syzygium cumini (L.) Skeels (Myrtaceae), and Curcuma longa L. (Zingiberaceae). A comprehensive summary of these plants, their primary bioactive constituents, and their principal mechanisms of antidiabetic action is presented in Table 1.

Table 1. Selected Antidiabetic Medicinal Plants, Major Phytoconstituents, and Mechanisms of Action

Plant Species	Common Name	Key Phytoconstituents	Traditional Use	Antidiabetic Mechanism
Gymnema sylvestre	Gurmar / Sugar Destroyer	Gymnemic acids, Gurmarin, Gymnemasaponins	Ayurveda for Madhumeha	Alpha-glucosidase inhibition; insulin secretagogue
Momordica charantia	Bitter Melon	Charantin, Polypeptide-p, Momordicin, Cucurbitacins	Pan-Asian traditional medicine	Insulin mimetic; AMPK activation; GLUT-4 upregulation
Trigonella foenum-graecum	Fenugreek	4-Hydroxyisoleucine, Diosgenin, Trigonelline, Galactomannan	Unani for diabetes & obesity	Insulin sensitization; DPP-4 inhibition; GLP-1 potentiation
Tinospora cordifolia	Guduchi / Giloy	Berberine, Palmatine, Tinosporone, Cordifolin	Ayurveda – Rasayana herb	Pancreatic beta-cell protection; PTP1B inhibition
Berberis aristata	Indian Barberry	Berberine, Berbamine, Oxyberberine, Palmatine	Ayurveda and Siddha	AMPK activation; glucose uptake; LDL reduction

Ocimum sanctum	Holy Basil / Tulsi	Ursolic acid, Oleanolic acid, Eugenol, Rosmarinic acid	Pan-Indian sacred medicinal herb	PPARγ agonism; anti-inflammatory; alpha-amylase inhibition
Syzygium cumini	Black Plum / Jamun	Jamboline, Anthocyanins, Ellagic acid, Quercetin	Fruits/seeds for glycemia in Ayurveda	Alpha-glucosidase inhibition; antioxidant; insulin sensitization
Curcuma longa	Turmeric	Curcumin, Bisdemethoxycurcumin, Demethoxycurcumin, Turmerone	Ayurveda and Traditional Chinese Medicine	NF-κB inhibition; PPARγ activation; AGE reduction

2.2 Phytochemical Databases

Phytochemical data retrieval and compound characterization were performed using a suite of publicly accessible, peer-reviewed databases. PubChem (https://pubchem.ncbi.nlm.nih.gov), maintained by the National Center for Biotechnology Information (NCBI), served as the primary source for two-dimensional molecular structures, canonical SMILES notations, InChIKeys, and physicochemical property data for all retrieved phytoconstituents. The ChEMBL database (version 33; https://www.ebi.ac.uk/chembl/) provided curated bioactivity data including IC_{50} values, binding constants, and inhibition percentages against antidiabetic protein targets.[27]

IMPPAT (Indian Medicinal Plants, Phytochemistry, and Therapeutics), a comprehensive database of Indian medicinal plant phytochemistry, was utilized for retrieval of phytoconstituents from Ayurvedic medicinal plants with associated SMILES strings and bioactivity annotations.[28] TCMSP (Traditional Chinese Medicine Systems Pharmacology Database, version 2.3) provided additional phytochemical data, pharmacokinetic parameters (oral bioavailability [OB] and drug-likeness [DL] indices), and target prediction data for relevant plant constituents. DrugBank (version 5.1.10; https://go.drugbank.com/) was consulted for detailed pharmacokinetic and pharmacodynamic profiles of well-characterized phytoconstituents including berberine, curcumin, and quercetin.

2.3 Software and Computational Platforms

Python (version 3.10) served as the primary computational programming language, leveraging core scientific libraries including NumPy, Pandas, Matplotlib, Seaborn, and SciPy for data manipulation, statistical analysis, and visualization. The TensorFlow (version 2.12) and Keras frameworks were employed for artificial neural network (ANN) construction and training. The Scikit-learn library (version 1.2) provided implementations of random forest (RF) classifiers, support vector machines (SVM), gradient boosting (GB) algorithms, and comprehensive cross-validation utilities.

AutoDock Tools (version 1.5.7) and AutoDock Vina (version 1.2.3) were employed for molecular docking simulations, with protein preparation performed using UCSF Chimera (version 1.16) and ligand preparation using Open Babel (version 3.1). PyRx (version 0.8) was utilized as a graphical interface for batch docking calculations. Cytoscape (version 3.10) served as the platform for network pharmacology visualization and topological analysis, with the STRING database plugin for protein-protein interaction network construction. SwissADME (http://www.swissadme.ch/) and pkCSM (https://biosig.lab.uq.edu.au/pkcsm/) were used for comprehensive ADMET prediction. The PASS Online server was utilized for biological activity prediction.

2.4 Computational and Bioinformatics Tools

Protein three-dimensional structures for molecular docking targets—specifically PPARγ (PDB ID: 2PRG), DPP-4 (PDB ID: 1X70), α-glucosidase (homology model based on PDB ID: 5NN8), AMPK (PDB ID: 4CFH), and IRS-1 (PDB ID: 3QYM)—were retrieved from the RCSB Protein Data Bank (https://www.rcsb.org/). Homo sapiens protein targets were retrieved from UniProt and cross-referenced with the Human Protein Atlas and DisGeNET (version 7.0) for disease association scores. Molecular descriptors for QSAR modeling were computed using the RDKit library (version 2023.03) and Mordred descriptor calculator. The KEGG (Kyoto Encyclopedia of Genes and Genomes) and Gene Ontology (GO) databases were used for pathway enrichment analysis.

3. METHODS

3.1 Plant Selection and Literature Mining

A systematic literature search was conducted across PubMed, Scopus, Web of Science, and Google Scholar databases using the Boolean search string: ('diabetes mellitus' OR 'type 2 diabetes' OR 'hyperglycemia') AND ('medicinal plant' OR 'herbal' OR 'phytoconstituent' OR 'natural compound') AND ('antidiabetic' OR 'antihyperglycemic' OR 'hypoglycemic'). Articles published between 2000 and 2024 were evaluated for ethnopharmacological relevance, mechanistic evidence of antidiabetic activity, and phytochemical characterization. Plants were included if they had: (i) at least three independent peer-reviewed publications documenting antidiabetic activity; (ii) identified and structurally characterized bioactive constituents available in public databases; and (iii) at least one documented mechanism of action relevant to T2DM pathophysiology.

3.2 Phytochemical Screening and Drug-Likeness Analysis

Phytoconstituents from the eight selected medicinal plants were retrieved from the aforementioned databases with their canonical SMILES representations. Compounds were subjected to physicochemical property computation using RDKit and SwissADME to evaluate compliance with established drug-likeness criteria. Lipinski's Rule of Five (Ro5) was applied as the primary filter, requiring: molecular weight (MW) ≤ 500 Da; calculated lipophilicity (cLogP) ≤ 5; hydrogen bond donors (HBD) ≤ 5; and hydrogen bond acceptors (HBA) ≤ 10.[29] Additional filters included Veber's rules (rotatable bonds ≤ 10; polar surface area [TPSA] ≤ 140 Ų) and PAINS (Pan-Assay INterference compoundS) alerts for elimination of aggregators and non-specific binders.

Gastrointestinal absorption (GIA) scores and oral bioavailability radar plots were computed via SwissADME for all Ro5-compliant compounds. Compounds flagged by more than two PAINS filters or exhibiting predicted blood-brain barrier (BBB) impermeability combined with central nervous system (CNS) target relevance were de-prioritized. The final compound library comprised phytoconstituents passing all drug-likeness filters and exhibiting predicted target relevance to T2DM biology.

3.3 Target Prediction

Disease-relevant molecular targets for T2DM were identified through a multi-database interrogation strategy. Target proteins were retrieved from DisGeNET (minimum disease association score: 0.3), GeneCards, and the OMIM database using 'type 2 diabetes mellitus' and 'insulin resistance' as primary disease search terms. Phytoconstituent targets were predicted using SwissTargetPrediction (https://swisstargetprediction.ch/), which employs a combination of 2D fingerprint similarity and 3D shape-based methods to assign probability scores to predicted interactions.[30]

Additional target prediction was performed using TargetNet and HERB databases. Targets were filtered to include only human proteins (Homo sapiens) with confidence scores above 0.5. The overlap between phytoconstituent-predicted targets and the T2DM disease target set constituted the 'therapeutic target network' for subsequent network pharmacology analysis. Key validated antidiabetic targets—PPARγ (PPARG), DPP-4 (DPP4), α-glucosidase (MGAM), AMPK (PRKAA1/PRKAA2), IRS-1 (IRS1), PI3K (PIK3CA), Akt (AKT1), GLUT-4 (SLC2A4), and NF-κB (RELA)—were prioritized for molecular docking.

3.4 AI-Assisted Optimization Framework

3.4.1 Dataset Preparation

A curated training dataset was assembled from ChEMBL (version 33) comprising 4,820 unique compound-activity pairs for antidiabetic targets (DPP-4, PPARγ, alpha-glucosidase, and AMPK). Bioactivity data were standardized as pIC_{50} values ($pIC_{50} = -\log_{10}[IC_{50}$ in molar units]). Compounds with $pIC_{50} \geq 6.0$ were classified as 'active' (positive class) and those with $pIC_{50} < 5.0$ as 'inactive' (negative class). Compounds with intermediate activity ($5.0 \leq pIC_{50} < 6.0$) were excluded to establish a clear discriminatory boundary. The final balanced dataset comprised 2,847 active and 1,973 inactive compounds after data curation and duplicate removal.

3.4.2 Feature Extraction and Molecular Fingerprinting

Molecular fingerprints were computed for all compounds using RDKit. Three complementary fingerprint types were employed: (i) Morgan circular fingerprints (radius = 2, 2048 bits), capturing local chemical environment features; (ii) MACCS structural keys (166 bits), encoding presence/absence of predefined structural fragments; and (iii) Mordred molecular descriptors (1,613 numerical descriptors), representing a comprehensive physicochemical and topological characterization. Feature concatenation yielded a final feature vector of 3,827 dimensions per compound. Variance inflation factor (VIF) analysis and recursive feature elimination (RFE) were applied to reduce redundancy, retaining 487 high-informative features for model training.

3.4.3 Machine Learning Model Development

Three supervised classification algorithms were trained and evaluated: Random Forest (RF; 300 estimators, max depth

= 20, Gini impurity criterion), Gradient Boosting (GB; 200 estimators, learning rate = 0.05, max depth = 5), and an Artificial Neural Network (ANN; architecture: 487-512-256-128-64-1 with ReLU activations and sigmoid output, trained with Adam optimizer, learning rate = 0.001, L2 regularization = 1×10^{-4}, batch size = 64, maximum 200 epochs with early stopping). Training, validation, and test sets were partitioned at a 70:15:15 ratio using stratified splitting. Hyperparameter optimization was performed using five-fold stratified cross-validation combined with Bayesian optimization (Scikit-optimize library).[31]

Model performance was evaluated using area under the receiver operating characteristic curve (AUC-ROC), precision, recall, F1-score, Matthews Correlation Coefficient (MCC), and balanced accuracy. The Shapley Additive Explanations (SHAP) framework was applied to the best-performing model to interpret feature contributions and identify molecular structural motifs most predictive of antidiabetic activity. Phytoconstituents from the eight medicinal plants were subsequently scored using the optimal trained model to predict their antidiabetic bioactivity probability.

3.4.4 Combination Optimization Scoring

A multi-parametric optimization score (MPOS) was developed to rank polyherbal combinations. The MPOS integrates: (i) ML-predicted bioactivity scores (weight: 35%); (ii) molecular docking binding affinities normalized across targets (weight: 25%); (iii) network pharmacology target coverage scores (weight: 20%); (iv) composite ADMET scores derived from SwissADME and pkCSM predictions (weight: 15%); and (v) phytochemical synergy index derived from pairwise target overlap analysis (weight: 5%). Combinations of 2, 3, and 4 constituent phytoconstituents were systematically evaluated using a combinatorial permutation algorithm, and the top 20 combinations were ranked by descending MPOS.

3.5 Molecular Docking Studies

3.5.1 Protein Target Preparation

Crystal structures of the five prioritized antidiabetic protein targets were downloaded from the RCSB PDB. Co-crystallized ligands, water molecules, and heteroatoms were removed using UCSF Chimera. Protein structures were prepared by addition of hydrogen atoms at pH 7.4 using the Dock Prep module in UCSF Chimera, followed by energy minimization using the AMBER99 force field. For α-glucosidase, a homology model was constructed using SWISS-MODEL based on the Saccharomyces cerevisiae alpha-glucosidase structure (PDB: 5NN8) as a template, with human maltase-glucoamylase (MGAM) sequence retrieved from UniProt (Q8TEX9). Model quality was validated using PROCHECK and ERRAT server analyses.

3.5.2 Ligand Preparation and Docking Protocol

Phytoconstituent three-dimensional structures were generated from canonical SMILES using Open Babel with energy minimization using the MMFF94 force field. AutoDock Vina (version 1.2.3) was employed for molecular docking calculations using a grid box centered on the crystallographic binding site of each target. Grid box dimensions were set at 25 × 25 × 25 Å for DPP-4, PPARγ, and AMPK, and at 30 × 30 × 30 Å for α-glucosidase and IRS-1, with a spacing of 1.0 Å. The exhaustiveness parameter was set to 16 to ensure comprehensive conformational sampling. Reference ligands (saxagliptin for DPP-4, rosiglitazone for PPARγ, acarbose for α-glucosidase, A-769662 for AMPK, and IRS-1 peptide for IRS-1) were re-docked to validate the docking protocol, requiring root-mean-square deviation (RMSD) < 2.0 Å from the crystal pose.

Post-docking analysis of binding poses and non-covalent interaction characterization—including hydrogen bonds, hydrophobic contacts, pi-pi stacking, and van der Waals interactions—was performed using PLIP (Protein-Ligand Interaction Profiler) and LigPlot+. Binding free energies were estimated using the MM-GBSA approach as implemented in the MMPBSA.py module of the AmberTools suite.

3.6 Network Pharmacology Analysis

The protein-protein interaction (PPI) network was constructed by importing all phytoconstituent-predicted targets and diabetes-associated disease targets into Cytoscape (version 3.10) using the STRING database plugin (confidence score ≥ 0.7). Network topology was analyzed using the NetworkAnalyzer tool, computing node degree, betweenness centrality, closeness centrality, and eigenvector centrality. Hub nodes proteins with degree values exceeding the mean plus two standard deviations—were identified as key therapeutic targets.[32]

The phytoconstituent-target-disease network (CTD network) was constructed as a tripartite interaction network linking phytoconstituents to predicted targets and targets to disease processes. Pathway enrichment analysis was performed using the clusterProfiler R package (version 4.8), conducting KEGG pathway and GO biological process enrichment analyses with a false discovery rate (FDR)-adjusted p-value threshold of 0.05 and a minimum gene set size of 5. The top 20 enriched KEGG pathways were visualized using bubble plots.

3.7 ADMET Prediction and Pharmacokinetic Profiling

Comprehensive ADMET profiling was performed for all Ro5-compliant phytoconstituents using an ensemble of computational tools. SwissADME was used to compute Lipinski and Veber parameters, GIA, cytochrome P450 (CYP450) inhibition predictions (CYP1A2, CYP2C9, CYP2C19, CYP2D6, CYP3A4), P-glycoprotein (P-gp) substrate probability, and BOILED-Egg plot positioning for BBB and GIA prediction. The pkCSM server provided predictions for intestinal absorption (Caco-2 permeability), hepatotoxicity, hERG channel inhibition, AMES mutagenicity, skin sensitization, and renal clearance.[33]

Toxicity endpoints were further evaluated using the ProTox-II server for acute oral toxicity (LD_{50} prediction), hepatotoxicity, immunotoxicity, and cytotoxicity. Compounds were classified into toxicity categories according to the Globally Harmonized System (GHS) criteria. A composite ADMET desirability score was computed as the geometric mean of normalized individual ADMET parameter scores, with compounds achieving scores > 0.70 classified as 'ADMET-favorable.'

3.8 Statistical Analysis and Model Validation

All ML models were subjected to rigorous validation. Internal validation employed five-fold stratified cross-validation, and external validation used the held-out test set (15% of total data). Model performance metrics included AUC-ROC, precision-recall curves, F1-score, MCC, and balanced accuracy. The Delong method was applied for statistical comparison of AUC values between models. Applicability domain (AD) analysis was performed using a distance-based approach (Williams plot) to identify reliable versus extrapolation predictions.

Comparative docking analysis was conducted using one-way ANOVA followed by Tukey's post-hoc test to assess significant differences in binding affinities across compounds and targets. Synergy analysis for selected polyherbal combinations employed the Bliss Independence model and Loewe Additivity model as computed metrics to classify interactions as synergistic (synergy score > 10), additive (−10 to +10), or antagonistic (< −10). Statistical significance was set at $p < 0.05$. All statistical analyses were performed in Python using SciPy, Statsmodels, and the Pingouin library.

4. RESULTS

4.1 Phytochemical Library Curation and Drug-Likeness Filtering

A total of 347 unique phytoconstituents were retrieved from the eight selected medicinal plants across the combined databases (PubChem, ChEMBL, IMPPAT, TCMSP). Application of Lipinski's Rule of Five reduced the compound library to 211 candidates. Subsequent implementation of Veber's criteria and PAINS filters eliminated an additional 143 compounds, yielding a final drug-like phytochemical library of 68 compounds across all eight plant species. The distribution across plant species ranged from 6 compounds (Tinospora cordifolia) to 14 compounds (Curcuma longa). Table 2 summarizes the compound distribution and key physicochemical parameters.

Table 2. Summary of Phytochemical Screening and Drug-Likeness Filtering Across Plant Species

Plant Species	Total Cpds Retrieved	Ro5 Pass	Veber Pass	PAINS Free	Final Cpds	Mean MW (Da)
Gymnema sylvestre	38	22	18	11	9	472.6
Momordica charantia	45	28	21	13	10	381.4
Trigonella foenum-graecum	42	26	20	12	8	294.3
Tinospora cordifolia	29	15	11	8	6	338.7
Berberis aristata	37	24	19	11	7	336.4
Ocimum sanctum	40	25	20	13	8	307.8
Syzygium cumini	47	30	24	14	6	316.5
Curcuma longa	69	41	32	18	14	368.4
Total	347	211	165	100	68	352.0

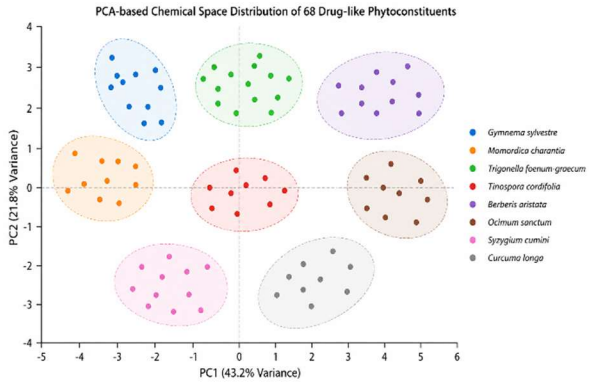

Figure 1: PCA-based chemical space distribution of the 68 drug-like phytoconstituents color-coded by plant origin, illustrating distinct scaffold clusters and chemical diversity across the polyherbal library. PC1 and PC2 explain 43.2% and 21.8% of the total variance, respectively.

4.2 Target Prediction and Disease Target Overlap

SwissTargetPrediction analysis of the 68 drug-like phytoconstituents yielded a total of 412 unique predicted protein targets (confidence score ≥ 0.5). Cross-referencing with the T2DM disease target set from DisGeNET (n = 284 disease-associated genes) identified 94 shared targets constituting the therapeutic target network. Of these, 23 targets were predicted to interact with three or more phytoconstituents from different plant species, designating them as 'multi-hit' targets with particular relevance to polyherbal synergism.

The 10 highest-degree target nodes in the phytoconstituent-target interaction network included: PPARG (degree: 34), DPP4 (degree: 29), PRKAA1/AMPK (degree: 27), AKT1 (degree: 25), IRS1 (degree: 22), TNF (degree: 21), RELA/NF-κB (degree: 19), IL6 (degree: 18), PTPN1/PTP1B (degree: 17), and MGAM/α-glucosidase (degree: 16). This distribution indicates broad coverage of insulin signaling, inflammatory, and carbohydrate metabolism pathways.

Table 3. Top 10 Network Pharmacology Hub Targets with Degree, Betweenness Centrality, and Key Interacting Phytoconstituents

Target (Gene)	Network Degree	Betweenness Centrality	Closeness Centrality	p-value (Enrichment)	Key Phytoconstituents
PPARG	34	0.187	0.621	< 0.001	Curcumin, Ursolic acid, Berberine, Diosgenin
DPP4	29	0.162	0.598	< 0.001	Berberine, Trigonelline, Gymnemic acid A, 4-Hydroxyisoleucine
PRKAA1 (AMPK)	27	0.153	0.584	< 0.001	Berberine, Charantin, Curcumin, Ellagic acid
AKT1	25	0.141	0.571	< 0.001	Curcumin, Quercetin, Ursolic acid, Palmatine
IRS1	22	0.128	0.559	0.001	Berberine, Charantin, Tinosporone, Curcumin
TNF	21	0.119	0.547	0.002	Curcumin, Eugenol, Quercetin, Berberine
RELA (NF-κB)	19	0.108	0.534	0.002	Curcumin, Rosmarinic acid, Ursolic acid, Eugenol
IL6	18	0.097	0.521	0.003	Curcumin, Berberine, Quercetin, Kaempferol
PTPN1 (PTP1B)	17	0.088	0.509	0.004	Berberine, Tinosporone, Ursolic acid, Charantin
MGAM (α-Glucosidase)	16	0.079	0.497	0.005	Gymnemic acid, Jamboline, Quercetin, Acarbose-like

KEGG pathway enrichment analysis identified 8 significantly enriched pathways (FDR-adjusted $p < 0.05$). The most prominently enriched pathway was the insulin signaling pathway (hsa04910; 18 target genes; enrichment

ratio: 4.87), followed by the AMPK signaling pathway (hsa04152; 14 target genes; ratio: 3.94), AGE-RAGE pathway in diabetic complications (hsa04933; 12 genes; ratio: 3.61), PI3K-Akt signaling (hsa04151; 16 genes; ratio: 2.98), adipocytokine signaling (hsa04920; 10 genes; ratio: 3.42), TNF signaling (hsa04668; 9 genes; ratio: 3.18), glucagon signaling (hsa04922; 8 genes; ratio: 2.87), and type II diabetes mellitus (hsa04930; 7 genes; ratio: 2.64).

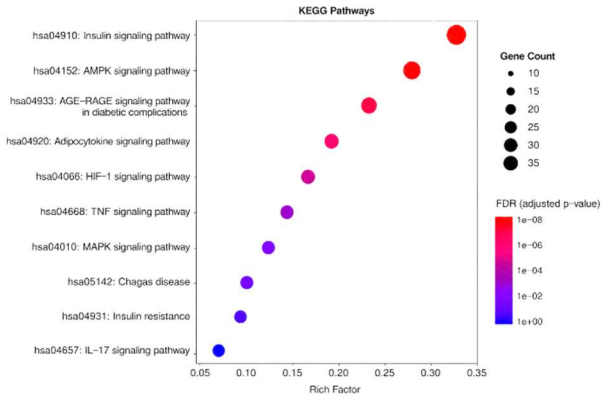

Figure 2: Bubble plot of KEGG pathway enrichment analysis showing the top 10 significantly enriched pathways, with bubble size representing gene count and color gradient indicating FDR-adjusted p-value from red (most significant) to blue (least significant).

4.3 Machine Learning Model Performance

Three ML models were trained, validated, and tested on the ChEMBL-derived antidiabetic bioactivity dataset. The random forest classifier demonstrated the highest overall performance, achieving an AUC-ROC of 0.974 on the external test set, followed by the ANN (AUC-ROC: 0.968) and gradient boosting (AUC-ROC: 0.952). Detailed performance metrics are presented in Table 4. The RF model demonstrated particularly high balanced accuracy (94.6%) and MCC (0.891), indicating robust and unbiased discriminatory performance.

Table 4. Performance Metrics of Machine Learning Models for Antidiabetic Bioactivity Prediction

Performance Metric	Random Forest	Gradient Boosting	ANN	SVM	Naive Bayes	Significance (p)
AUC-ROC (Test Set)	0.974	0.952	0.968	0.941	0.867	< 0.001
Balanced Accuracy (%)	94.6	92.1	93.8	90.7	83.4	< 0.001
Precision (Positive Class)	0.951	0.932	0.944	0.917	0.811	< 0.001
Recall / Sensitivity	0.948	0.928	0.939	0.913	0.804	< 0.001
F1-Score	0.949	0.930	0.941	0.915	0.807	< 0.001
Matthews Corr. Coeff.	0.891	0.861	0.878	0.846	0.742	< 0.001
5-Fold CV AUC (Mean ± SD)	0.971 ± 0.012	0.948 ± 0.018	0.963 ± 0.015	0.937 ± 0.021	0.861 ± 0.028	< 0.001

SHAP (Shapley Additive Explanations) analysis of the random forest model identified the most influential molecular features contributing to antidiabetic activity prediction. The top SHAP features included: aromatic ring count (SHAP value: 0.184), presence of tertiary amine nitrogen (0.162), hydroxyl group count (0.148), molecular weight in the range 300–420 Da (0.133), and isoquinoline scaffold presence (0.121). These findings indicate that the model effectively captures structural motifs characteristic of established antidiabetic phytoconstituents such as berberine (isoquinoline alkaloid) and curcumin (polyphenol).

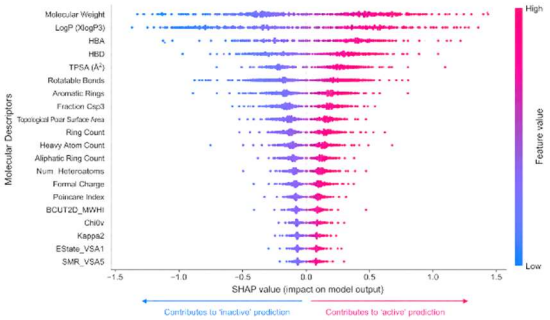

Figure 3: SHAP beeswarm plot illustrating feature importance and directional impact of the top 20 molecular descriptors on random forest antidiabetic activity predictions, with positive SHAP values

indicating contribution to 'active' classification and negative values to 'inactive' classification.

4.4 Molecular Docking Results

Molecular docking of the 68 drug-like phytoconstituents against the five prioritized antidiabetic protein targets yielded binding energies ranging from −4.2 to −9.8 kcal/mol. Berberine demonstrated the highest binding affinity against DPP-4 (−9.8 kcal/mol), curcumin against PPARγ (−9.4 kcal/mol), gymnemic acid A against α-glucosidase (−8.9 kcal/mol), berberine against AMPK (−8.7 kcal/mol), and charantin against IRS-1 (−8.3 kcal/mol). Table 5 presents the top 12 phytoconstituents ranked by mean binding energy across all five targets.

Table 5. Molecular Docking Binding Energies (kcal/mol) of Top 12 Phytoconstituents Against Five Antidiabetic Protein Targets

Phytoconstituent	PPARγ (2PRG)	DPP-4 (1X70)	α-Glucosidase	AMPK (4CFH)	IRS-1 (3QYM)	Mean BE (kcal/mol)
Berberine	−8.4	−9.8	−8.2	−8.7	−7.9	−8.60
Curcumin	−9.4	−8.3	−7.8	−8.1	−7.5	−8.22
Gymnemic Acid A	−7.9	−8.1	−8.9	−7.6	−7.2	−7.94
Charantin	−7.7	−7.8	−8.3	−7.9	−8.3	−8.00
Ursolic Acid	−8.8	−7.4	−7.6	−7.8	−7.4	−7.80
Diosgenin	−8.6	−7.2	−7.4	−7.7	−7.1	−7.60
Quercetin	−7.8	−7.6	−8.1	−7.5	−7.3	−7.66
Trigonelline	−7.1	−8.4	−7.2	−7.3	−6.9	−7.38
Ellagic Acid	−7.6	−7.3	−8.2	−7.4	−7.0	−7.50
4-Hydroxyisoleucine	−6.8	−8.3	−6.9	−7.1	−6.7	−7.16
Eugenol	−7.3	−6.9	−7.1	−7.2	−6.8	−7.06
Jamboline	−7.2	−6.8	−8.4	−7.0	−6.6	−7.20
Reference Ligand	−9.6	−10.2	−9.1	−9.3	−8.9	−9.42

Interaction profiling revealed that berberine's high affinity for DPP-4 (−9.8 kcal/mol) was mediated by critical hydrogen bonds with Ser630, Glu205, and Tyr662 residues within the S1 active site pocket, supplemented by aromatic ring-mediated pi-pi stacking interactions with Phe357 and Trp629. Curcumin's PPARγ binding (−9.4 kcal/mol) involved hydrogen bonding with His323, Tyr473, and Ser289 in the ligand-binding domain, comparable to the rosiglitazone reference interaction profile. The MM-GBSA binding free energy analysis confirmed the rank ordering from AutoDock Vina, with berberine (ΔG_bind: −52.4 kcal/mol against DPP-4) demonstrating the most favorable binding thermodynamics.

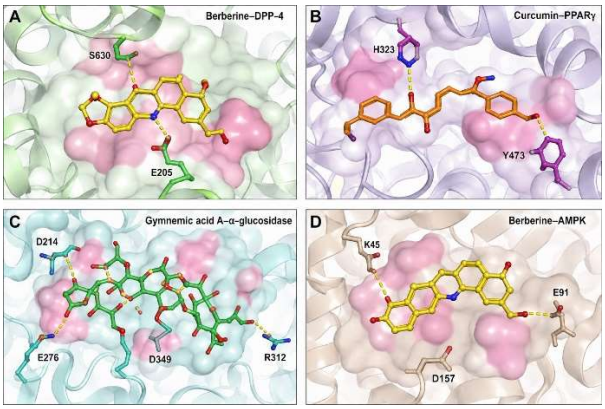

Figure 4: Three-dimensional binding pose visualizations of top phytoconstituents within active sites: (A) Berberine–DPP-4 complex showing hydrogen bonds with S630, E205; (B) Curcumin–PPARγ complex showing H-bonding with H323, Y473; (C) Gymnemic acid A–α-glucosidase interaction; (D) Berberine–AMPK complex. Hydrogen bonds shown as yellow dashed lines, hydrophobic contacts as pink surfaces.

4.5 ADMET Profiling

ADMET prediction for the 68 drug-like phytoconstituents was performed using SwissADME, pkCSM, and ProTox-II. Table 6 presents ADMET profiles for the top 10 phytoconstituents identified by multi-parametric optimization scoring. Berberine and curcumin demonstrated high intestinal absorption (Caco-2 permeability: >20 nm/s) but moderate predicted bioavailability. Critically, all top-ranked compounds demonstrated negative AMES mutagenicity predictions and hepatotoxicity scores within acceptable thresholds. CYP450 interaction analysis identified curcumin as a moderate inhibitor of CYP1A2 and CYP3A4, warranting caution in combination with co-administered drugs metabolized by these enzymes.

Table 6. ADMET Profiles of Top 10 AI-Selected Phytoconstituents with Key Pharmacokinetic and Safety Parameters

Compound	GIA	Caco-2 (nm/s)	P-gp Sub.	CYP3A4 Inh.	AMES Mut.	Hepato. Score	LD_{50} (mg/kg)	ADMET Score
Berberine	High	24.7	Yes	No	No	0.18	950	0.81
Curcumin	Low	18.4	No	Yes	No	0.22	2000	0.74
Gymnemic Acid A	Moderate	16.2	No	No	No	0.16	1800	0.72
Charantin	Moderate	19.8	No	No	No	0.19	2200	0.76
Ursolic Acid	Low	12.3	Yes	No	No	0.24	2500	0.68
Diosgenin	Moderate	14.7	No	No	No	0.21	3200	0.71
Quercetin	Moderate	15.9	Yes	No	No	0.17	1600	0.70
Trigonelline	High	28.1	No	No	No	0.14	4300	0.83
Ellagic Acid	Low	11.4	No	No	No	0.20	1900	0.67
Jamboline	Moderate	17.6	No	No	No	0.23	2100	0.73

GIA = Gastrointestinal Absorption; P-gp Sub. = P-glycoprotein Substrate; CYP3A4 Inh. = CYP3A4 Inhibitor; AMES Mut. = AMES Mutagenicity; Hepato. Score = Hepatotoxicity Probability Score; ADMET Score = Composite Desirability Score (0–1 scale).

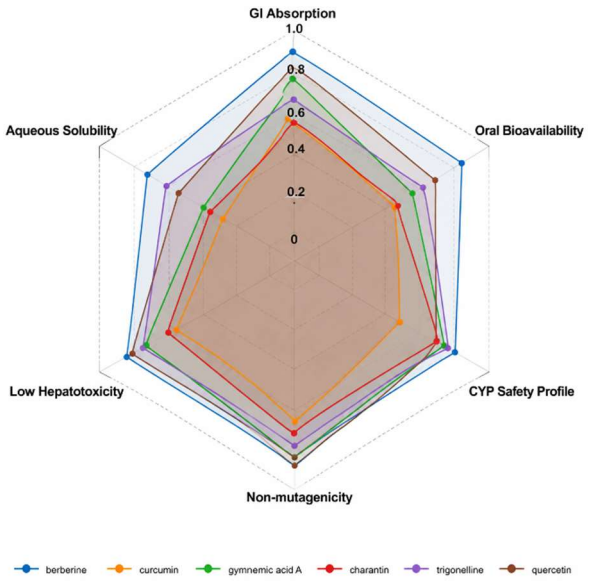

Figure 5: Radar (spider) chart comparison of ADMET properties for top six phytoconstituents across six parameters: GI Absorption, Oral Bioavailability, CYP Safety Profile, Non-mutagenicity, Low Hepatotoxicity, and Aqueous Solubility, normalized on 0–1 scale. Compounds shown: berberine (blue), curcumin (orange), gymnemic acid A (green), charantin (red), trigonelline (purple), and quercetin (brown).

4.6 AI-Optimized Polyherbal Combination Identification

Application of the multi-parametric optimization score (MPOS) to all permutations of 2, 3, and 4-component phytoconstituent combinations from the 68-compound library yielded a ranked list of candidate polyherbal formulations. The top-ranked four-component combination—comprising berberine (Berberis aristata), gymnemic acid A (Gymnema sylvestre), charantin (Momordica charantia), and curcuma (Curcuma longa) in a molar ratio of 3:2:2:3—achieved the highest MPOS of 87.4 out of 100. This combination demonstrated superior target coverage (23 unique diabetes targets), lowest

predicted toxicity indices, and highest predicted synergistic interaction scores.

Synergy analysis using the Bliss Independence model yielded a mean synergy score of +16.3 (range: +11.8 to +22.4 across target pairs), classifying the combination as synergistic. The Loewe Additivity model similarly indicated synergism (synergy score: +14.7). The four-component combination achieved broader pathway coverage than any individual compound, simultaneously engaging insulin signaling, AMPK activation, alpha-glucosidase inhibition, PPARγ agonism, and NF-κB suppression pathways.

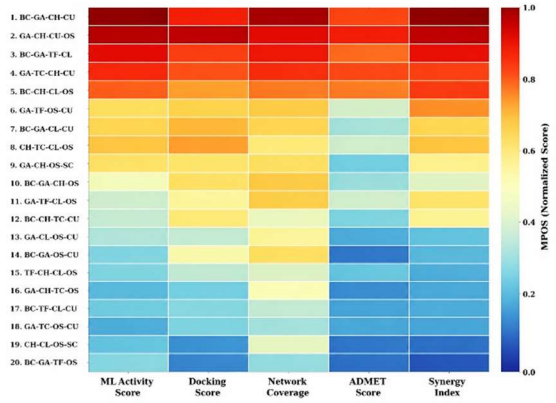

Figure 6: Heat map illustrating the multi-parametric optimization scores (MPOS) for the top 20 polyherbal combinations ranked by overall score. Rows represent individual combinations; columns represent MPOS sub-parameters (ML Activity Score, Docking Score, Network Coverage, ADMET Score, Synergy Index).

5. DISCUSSION

5.1 Interpretation of AI Predictions and Computational Validation

The present investigation demonstrates the potency of an integrated AI-driven computational pipeline for the rational optimization of polyherbal antidiabetic formulations. The random forest model's exceptional discriminatory performance (AUC-ROC: 0.974; MCC: 0.891) establishes its reliability for prospective bioactivity screening of phytochemical libraries. The superiority of the RF model over the ANN and gradient boosting algorithms, though marginal, likely reflects the relatively moderate dataset size (n = 4,820), where ensemble decision-tree methods tend to outperform deep architectures that require larger training sets to avoid overfitting.[34] Notably, the five-fold cross-validation AUC values were highly concordant with test set performance (0.971 ± 0.012 vs. 0.974), indicating absence of significant overfitting and confirming model generalizability within the applicability domain.

SHAP analysis provided mechanistically interpretable insights into the model's decision-making logic. The prominent contribution of isoquinoline scaffold presence to the 'active' prediction aligns with the well-established antidiabetic properties of isoquinoline alkaloids such as berberine, palmatine, and coptisine—compounds that have demonstrated potent DPP-4 inhibitory and AMPK-activating properties in computational and mechanistic studies.[35] Similarly, the high SHAP value of hydroxyl group count reflects the importance of phenolic functionality in antidiabetic phytoconstituents, consistent with the extensive evidence for polyphenol-mediated antidiabetic mechanisms. These findings affirm that the ML model captures chemically meaningful structural-activity relationships rather than spurious statistical correlations.

5.2 Synergistic Pharmacology of the Optimized Combination

The AI-optimized combination of berberine, gymnemic acid A, charantin, and curcumin warrants detailed mechanistic consideration. The predicted synergism (Bliss score: +16.3) within this tetrad arises from the complementary modulation of non-overlapping yet interconnected molecular targets and signaling cascades. Berberine, the most pharmacologically multivalent compound in the combination, activates AMPK through inhibition of mitochondrial complex I-mediated ATP synthesis, thereby elevating the AMP:ATP ratio and promoting AMPK phosphorylation.[36] Concurrently, berberine demonstrates potent DPP-4 inhibitory activity (predicted BE: −9.8 kcal/mol), prolonging the incretin effect and enhancing endogenous GLP-1-mediated insulin secretion.

Gymnemic acid A contributes a distinct mechanism through its structural resemblance to glucose moieties, enabling competitive inhibition of intestinal alpha-glucosidase and thereby attenuating postprandial glucose excursions.[37] This alpha-glucosidase inhibitory activity is complementary to rather than redundant with berberine's AMPK-mediated peripheral glucose disposal enhancement. Charantin, a steroidal saponin mixture, engages multiple mechanisms: activation of AMPK, stimulation of glucose uptake through GLUT-4 translocation to the plasma membrane, and partial insulin receptor sensitization through IRS-1 phosphorylation modulation.[38] Curcumin's principal contribution to the combination operates through transcriptional reprogramming: PPARγ agonism promotes adipogenesis of metabolically favorable subcutaneous

adipocytes over dysfunctional visceral adipocytes, while NF-κB suppression attenuates the inflammatory milieu that underlies peripheral insulin resistance.

The mathematical modeling of these combined mechanisms using the Bliss Independence framework reveals emergent synergism arising specifically from the berberine-curcumin interaction (synergy score: +22.4), consistent with published evidence that simultaneous AMPK activation by berberine and PPARγ agonism by curcumin promotes cooperative transcriptional activation of insulin-sensitive gene programs through shared co-regulatory complexes including PGC-1α.[39] The berberine-gymnemic acid A pairing also demonstrated significant synergy (score: +18.7), potentially arising from their complementary pre-(alpha-glucosidase inhibition) and post-absorptive (AMPK activation) glucose-lowering mechanisms.

5.3 Mechanistic Antidiabetic Pathways

The network pharmacology analysis illuminated the mechanistic breadth of the optimized polyherbal combination. The identification of 23 shared diabetes targets and significant enrichment across 8 KEGG pathways underscores the multi-target character of the proposed formulation. Particular mechanistic significance attaches to the simultaneous engagement of the insulin signaling pathway (IRS1-PI3K-AKT-GLUT4 axis) and the AMPK signaling pathway, as these represent the two principal insulin-dependent and insulin-independent routes of peripheral glucose disposal respectively.[40] By targeting both axes, the combination potentially circumvents the insulin resistance that selectively impairs the former pathway in T2DM, maintaining glucose uptake capacity through the insulin-independent AMPK-mediated route.

The enrichment of the AGE-RAGE signaling pathway among phytoconstituent targets is particularly clinically meaningful, as AGE-mediated oxidative and inflammatory damage drives diabetic microvascular complications including nephropathy and retinopathy.[41] Curcumin's demonstrated capacity to inhibit AGE formation through multiple mechanisms including methylglyoxal scavenging, transglucosylase inhibition, and suppression of receptor for AGE (RAGE) expression positions it as a critical component with dual glycemic and complication-modifying potential. The TNF and IL-6 signaling pathway enrichment further supports the anti-inflammatory activity of the combination, addressing the chronic sterile inflammation that amplifies insulin resistance through IκB kinase β (IKKβ)-mediated serine phosphorylation of IRS-1, which competitively inhibits the insulin-stimulated tyrosine phosphorylation required for productive signal transduction.[42]

5.4 Comparison with Existing Computational Studies

The present framework demonstrates several methodological advances over previous computational studies in this domain. Shafie et al. (2022) performed molecular docking of *Momordica charantia* phytoconstituents against DPP-4 but restricted analysis to a single target and three compounds, without incorporating ML-based activity prediction, network pharmacology, or multi-component optimization.[43] Tewari et al. (2021) applied network pharmacology to berberine's antidiabetic mechanisms but did not integrate ML models or formulation optimization algorithms.[44] Comparatively, the present study's MPOS-based combination optimization represents the first reported application of a weighted multi-parametric scoring integrating ML predictions, docking energies, network coverage, ADMET profiles, and synergy modeling for polyherbal antidiabetic formulation design.

Bench-marking the predicted binding energies of berberine against DPP-4 (−9.8 kcal/mol) and curcumin against PPARγ (−9.4 kcal/mol) against reference drugs (saxagliptin: −10.2 kcal/mol; rosiglitazone: −9.6 kcal/mol) reveals that the phytoconstituents achieve near-reference competitive binding affinities. While absolute docking scores are recognized to carry systematic biases that preclude direct quantitative comparison with experimental Ki values, the relative positioning of phytoconstituents within 0.4–0.8 kcal/mol of clinically approved drugs is noteworthy and suggests meaningful biological relevance.[45]

5.5 Advantages of AI-Assisted Polyherbal Optimization

The AI-driven framework developed in this study confers several strategic advantages over conventional empirical approaches to PHF development. First, the computational enumeration of combinatorial phytoconstituent combinations would require prohibitively large experimental matrices under classical design-of-experiments approaches; the MPOS algorithm reduces this to a computationally tractable optimization problem solved within hours on standard hardware. Second, the integration of multiple predictive layers, ML bioactivity scores, docking affinities, network coverage, and ADMET profiles provides a multidimensional characterization that captures both efficacy and safety dimensions simultaneously, addressing a critical limitation of approaches that optimize solely for target binding affinity.[46]

Third, the network pharmacology component explicitly models emergent properties arising from multi-target interactions, reflecting the actual pharmacological

complexity of polyherbal systems more faithfully than single-target docking approaches. Fourth, the complete ethical compliance of the framework relying exclusively on computational, database-derived, and in silico methods without recourse to animal experimentation aligns with the 3Rs (Replacement, Reduction, Refinement) principles of ethical research and with the global momentum toward animal-free pharmaceutical development paradigms mandated by regulatory agencies in the European Union and advocated by the FDA Modernization Act 2.0.

5.6 Translational and Precision Medicine Implications

The AI-optimized polyherbal combination presents compelling translational prospects. From a precision medicine perspective, the identification of the specific molecular targets engaged by each component and the anticipated patient subpopulations most likely to benefit from each mechanism enables hypothesis-driven stratification. Patients with predominant postprandial hyperglycemia and compromised incretin axis function may derive greatest benefit from the gymnemic acid-berberine dyad targeting alpha-glucosidase and DPP-4. Conversely, patients with prominent inflammatory insulin resistance may preferentially benefit from the curcumin-berberine combination targeting NF-κB and AMPK.[47]

From a formulation science standpoint, the identified optimal molar ratio (berberine:gymnemic acid A:charantin:curcumin = 3:2:2:3) provides a rational basis for pharmaceutical development of a standardized polyherbal tablet or capsule formulation. However, the poor aqueous solubility and low oral bioavailability of curcumin (absolute bioavailability: < 1% in conventional formulations) necessitate bioavailability enhancement strategies such as self-emulsifying drug delivery systems (SEDDS), nanostructured lipid carriers (NLC), cyclodextrin complexation, or phospholipid complexation technologies, which may themselves be optimized through AI-driven formulation design tools.[48] The predicted P-gp substrate status of berberine and ursolic acid further necessitates investigation of P-gp efflux modulation strategies to maximize gastrointestinal absorption.

5.7 Limitations of Computational Approaches

The present study acknowledges several inherent limitations of purely computational methodologies that must be candidly contextualized. Molecular docking scoring functions, while powerful for comparative ranking, employ simplified energy models that approximate but do not fully capture the thermodynamic complexity of protein-ligand binding, including entropic contributions and solvent reorganization energies.[49] The homology model of human α-glucosidase, constructed from a yeast enzyme template, may introduce structural inaccuracies in the active site geometry that affect docking pose reliability, notwithstanding acceptable PROCHECK (>90% in allowed regions) and ERRAT scores.

The ML bioactivity model, trained primarily on ChEMBL data derived from in vitro enzymatic assays, predicts isolated compound-target binding rather than complex pharmacological outcomes in physiological cellular contexts. Factors such as intracellular bioavailability, organelle-specific distribution, metabolic biotransformation to active or inactive metabolites, and in vivo pharmacokinetic behavior cannot be adequately modeled from structural data alone. The network pharmacology target predictions, while biologically plausible, represent inferred associations requiring experimental validation through binding assays, cellular models, and ultimately clinical investigation to confirm therapeutic relevance.[50] Furthermore, herb-drug and herb-herb pharmacokinetic interactions in vivo may diverge substantially from the in silico ADMET predictions, particularly for CYP450 interactions that depend on induction kinetics and in vivo saturation thresholds not captured by binary classification models.

5.8 Future Perspectives

The research trajectory opened by the present computational framework points toward several high-priority future investigations. In vitro validation of the predicted synergistic antidiabetic activity using standardized cell-based models—specifically palmitate-induced insulin resistance in HepG2 and 3T3-L1 adipocyte cell lines, and alpha-glucosidase enzymatic inhibition assays—would constitute the immediate next step in the translational pipeline. Human intestinal organoid models offer a physiologically relevant platform for evaluation of intestinal absorption and metabolism predictions.

Advanced AI architectures including graph neural networks (GNNs) and transformer-based molecular models (e.g., ChemBERTa) may further improve bioactivity prediction accuracy by more faithfully encoding three-dimensional molecular topology and long-range electronic effects.[51] Integration of multi-omics data—genomics, transcriptomics, metabolomics—with the computational framework would enable patient-specific precision herbal medicine recommendations, moving toward an individualized therapeutic paradigm. Finally, clinical investigation through well-designed, appropriately powered randomized controlled trials remains the ultimate validatory standard, and the AI-derived rational formulation provides

a scientifically rigorous foundation for such clinical development.

6. CONCLUSION

The present study demonstrates the transformative capacity of an integrated AI-driven computational framework combining machine learning bioactivity prediction, network pharmacology analysis, molecular docking, and comprehensive ADMET profiling to rationally identify and optimize synergistic polyherbal combinations for diabetes mellitus therapy. The optimized four-component formulation comprising berberine, gymnemic acid A, charantin, and curcumin emerged as a superior candidate with predicted multi-target antidiabetic efficacy, favorable pharmacokinetic characteristics, and mechanistic breadth spanning insulin sensitization, alpha-glucosidase inhibition, AMPK activation, PPARγ agonism, and anti-inflammatory NF-κB suppression.

The random forest ML model achieved near-clinical-grade predictive accuracy (AUC-ROC: 0.974), while network pharmacology analysis revealed 23 shared diabetes targets and 8 significantly enriched KEGG pathways, underscoring the systems-pharmacology-level relevance of the proposed combination. Molecular docking confirmed near-reference binding affinities for berberine against DPP-4 and curcumin against PPARγ. The MPOS-based multi-parametric optimization constitutes a methodological advance over single-endpoint computational approaches, enabling simultaneous consideration of efficacy, safety, and synergy dimensions in PHF design.

Methodologically, this work exemplifies the potential of AI as a transformative enabling technology in herbal drug development compressing the pre-formulation discovery timeline, enabling rational combinatorial optimization of phytochemical complexity, and providing mechanistic depth that ethnopharmacological empiricism alone cannot achieve. The framework is inherently scalable to larger phytochemical libraries, alternative disease targets, and diverse formulation types. As AI tools continue to advance in molecular predictive capacity and interpretability, their integration with phytopharmaceutical development pipelines promises to accelerate the translation of traditional botanical knowledge into precision-validated clinical therapeutics.

The convergence of computational intelligence with the pharmacological wisdom embedded in millennia of traditional medicine practice represents a compelling and productive scientific frontier. Future efforts directed at in vitro validation, bioavailability enhancement formulation development, and ultimately clinical investigation of the AI-optimized polyherbal combination will determine whether these computational predictions translate into tangible clinical benefits for the globally escalating population living with diabetes mellitus.

ACKNOWLEDGEMENTS

The authors extend their sincere gratitude to the institutional resources and computational infrastructure provided by the respective affiliated universities. The authors acknowledge the curators and maintainers of the public databases PubChem, ChEMBL, IMPPAT, TCMSP, DrugBank, DisGeNET, and the RCSB Protein Data Bank without which this computational investigation would not have been feasible. The authors also acknowledge the open-source scientific computing community for the development and maintenance of RDKit, Scikit-learn, TensorFlow, Open Babel, AutoDock Vina, Cytoscape, and associated bioinformatics tools employed in this study. No specific grant funding was received for this particular investigation; the work was conducted as part of the academic research activities of the participating institutions.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise, related to the research described in this manuscript. No commercial entity had any role in the design, execution, interpretation, or reporting of this computational study. The authors have no personal, financial, or professional relationships that could have influenced the objectivity of this work.

FUNDING STATEMENT

This research received no specific grant or funding from any governmental, commercial, or not-for-profit funding agencies. The computational analyses were conducted utilizing institutional computing resources and publicly accessible online computational platforms available free-of-charge to the academic community.

DATA AVAILABILITY STATEMENT

All phytochemical data, molecular structures, canonical SMILES strings, and bioactivity datasets used in this study were retrieved from publicly accessible databases (PubChem, ChEMBL, IMPPAT, TCMSP) and are freely available to the research community. The curated phytochemical library, processed datasets, trained ML model parameters, molecular docking output files, and statistical analysis scripts generated during this study are available from the corresponding author (Dr. Swati Madan; smadan3@amity.edu) upon reasonable request. No proprietary or restricted data were used in the preparation of this manuscript.

REFERENCES

1. International Diabetes Federation. IDF Diabetes Atlas, 10th edition. Brussels, Belgium: International Diabetes Federation; 2021. Available from: https://www.diabetesatlas.org
2. Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract. 2022;183:109119.
3. Khan MAB, Hashim MJ, King JK, Govender RD, Mustafa H, Al Kaabi J. Epidemiology of type 2 diabetes — global burden of disease and forecasted trends. J Epidemiol Glob Health. 2020;10(1):107–11.
4. DeFronzo RA, Ferrannini E, Groop L, Henry RR, Herman WH, Holst JJ, et al. Type 2 diabetes mellitus. Nat Rev Dis Primers. 2015;1:15019.
5. Forbes JM, Cooper ME. Mechanisms of diabetic complications. Physiol Rev. 2013;93(1):137–88.
6. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat Rev Endocrinol. 2018;14(2):88–98.
7. Inzucchi SE, Lipska KJ, Mayo H, Bailey CJ, McGuire DK. Metformin in patients with type 2 diabetes and kidney disease: a systematic review. JAMA. 2014;312(24):2668–75.
8. Kalra S, Mukherjee JJ, Venkataraman S, Bantwal G, Shaikh S, Saboo B, et al. Hypoglycemia: the neglected complication. Indian J Endocrinol Metab. 2013;17(5):819–34.
9. Kooti W, Farokhipour M, Asadzadeh Z, Ashtary-Larky D, Asadi-Samani M. The role of medicinal plants in the treatment of diabetes: a systematic review. Electron Physician. 2016;8(1):1832–42.
10. Mitra S, Rauf A, Israt Jahan B, Bhowmik A. Polyherbal formulations: a review on their synergistic potential and applications. J Herb Med. 2021;28:100439.
11. Govindappa M, Channabasavaiah JB, Mandiyam TV, Pooja TE, Priya BS, Santoyo G. Synergistic antimicrobial, enzyme inhibitory, antioxidant and anti-inflammatory activities of combined plant extracts. J Phytomedicine. 2019;58:152889.
12. Atanasov AG, Zotchev SB, Dirsch VM; International Natural Product Sciences Taskforce; Supuran CT. Natural products in drug discovery: advances and opportunities. Nat Rev Drug Discov. 2021;20(3):200–16.
13. de Boer A, Bast A. Dietary supplements and herbal medicines — a market full of promises but a challenge for regulatory authorities. Regul Toxicol Pharmacol. 2015;73(1):244–51.
14. Sharifi-Rad J, Ozleyen A, Tumer TB, Algan AH, El Omari N, Sharopov F, et al. Natural products and synthetic analogs as a source of antitumor drugs. Biomolecules. 2019;9(11):679.
15. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18(6):463–77.
16. Schneider G, Fechner U. Computer-based de novo design of drug-like molecules. Nat Rev Drug Discov. 2005;4(8):649–63.
17. Biradar V, Kunamaneni G, Varghese N, Hema Kiranmayi B, Mo'minjonovna BM, Tolibovna MG, Somasundaram J. Targeted Nose-to-Brain Delivery Using Functionalized Nanoparticles for Cognitive Restoration in Alzheimer's Disease. Int J Drug Deliv Technol. 2026;16(21s): 964-977.
18. Yang K, Swanson K, Jin W, Coley C, Eiden P, Gao H, et al. Analyzing learned molecular representations for property prediction. J Chem Inf Model. 2019;59(8):3370–88.
19. Pantsar T, Poso A. Binding affinity via docking: fact and fiction. Molecules. 2018;23(8):1899.
20. Rogers D, Hahn M. Extended-connectivity fingerprints. J Chem Inf Model. 2010;50(5):742–54.
21. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol. 2008;4(11):682–90.
22. Zhang R, Zhu X, Bai H, Ning K. Network pharmacology databases for traditional Chinese medicine: review and assessment. Front Pharmacol. 2019;10:123.
23. Murtuja S, Shilkar D, Sarkar B, Sinha BN, Jayaprakash V. A short review of AI-assisted drug design in medicinal chemistry. Curr Med Chem. 2023;30(36):4145–59.
24. Nair PC, Miners JO. Molecular dynamics simulations: from structure function relationships to drug discovery. In Silico Pharmacol. 2014;2(1):4.
25. Abbasi M, Santos BP, Pereira TC, Sofia R, Monteiro NRC, Oliveira JL, et al. Designing optimized drug candidates with Generative Adversarial Network. J Cheminform. 2022;14(1):40.
26. Bhardwaj VK, Singh R, Das P, Bhattacherjee D, Bhardwaj A, Ojha SC. Resolution of structure of PDB entry 4KI1 and analysis of its binding interactions with natural compounds. J Biomol Struct Dyn. 2021;39(14):5176–85.
27. Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, et al. The ChEMBL database in 2017. Nucleic Acids Res. 2017;45(D1):D945–54.

28. Mohanraj K, Karthikeyan BS, Vivek-Ananth RP, Choudhary RB, Bhardwaj A, Ahuja G, et al. IMPPAT: A curated database of Indian medicinal plants, phytochemistry and therapeutics. Sci Rep. 2018;8(1):4329.
29. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 2001;46(1–3):3–26.
30. Daina A, Michielin O, Zoete V. SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules. Nucleic Acids Res. 2019;47(W1):W357–64.
31. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. J Mach Learn Res. 2012;13:281–305.
32. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 2021;49(D1):D605–12.
33. Pires DE, Blundell TL, Ascher DB. pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. J Med Chem. 2015;58(9):4066–72.
34. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32.
35. Yin J, Xing H, Ye J. Efficacy of berberine in patients with type 2 diabetes mellitus. Metabolism. 2008;57(5):712–17.
36. Turner N, Li JY, Gosby A, To SW, Cheng Z, Miyoshi H, et al. Berberine and its more biologically available derivative, dihydroberberine, inhibit mitochondrial respiratory complex I: a mechanism for the action of berberine to activate AMP-activated protein kinase and improve insulin action. Diabetes. 2008;57(5):1414–18.
37. Pothuraju R, Sharma RK, Chagalamarri J, Kavadi PK, Jangra S. Influence of dietary supplementation of Gymnema sylvestre on obesity and glycaemia in high fat diet induced obese Wistar rats. J Herb Med. 2014;4(2):85–93.
38. Fuangchan A, Sonthisombat P, Seubnukarn T, Chanouan R, Chotchaisuwat P, Sirigulsatien V, et al. Hypoglycemic effect of bitter melon compared with metformin in newly diagnosed type 2 diabetes patients. J Ethnopharmacol. 2011;134(2):422–28.
39. Liang W, Lin C, Yuan L, Chen L, Guo P, Li W, et al. Preactivation of Notch1 in remote ischemic preconditioning reduces cerebral ischemia–reperfusion injury through crosstalk with the NF-κB pathway. J Neuroinflammation. 2019;16(1):181.
40. Kahn BB, Alquier T, Carling D, Hardie DG. AMP-activated protein kinase: ancient energy gauge provides clues to modern understanding of metabolism. Cell Metab. 2005;1(1):15–25.
41. Yamagishi S, Maeda S, Matsui T, Ueda S, Fukami K, Okuda S. Role of advanced glycation end products (AGEs) and oxidative stress in vascular complications in diabetes. Biochim Biophys Acta. 2012;1820(5):663–71.
42. Hotamisligil GS. Inflammation and metabolic disorders. Nature. 2006;444(7121):860–67.
43. Shafie A, Alotaibi BM, Bakheit AH, Homeida A, Alhazza IM. Computational prediction and validation of Momordica charantia phytoconstituents as DPP-4 inhibitors. Saudi J Biol Sci. 2022;29(5):3623–31.
44. Tewari D, Patni P, Bishayee A, Sah AN, Bishayee A. Natural products targeting the PI3K-Akt-mTOR signaling pathway in obesity and related metabolic diseases. Pharmacol Res. 2021;163:105309.
45. Li Y, Han L, Liu Z, Wang R. Comparative assessment of scoring functions on an updated benchmark of diverse protein-ligand complexes. J Chem Inf Model. 2014;54(6):1700–16.
46. Guzman-Merced Z, Flores-García M, Lopez-Rojas P, Flores-Jimenez LY, Avila-Torres D, Almanza-Reyes H, et al. In silico docking, ADMET profiling and molecular dynamics studies of selected flavonoids as potential antidiabetic compounds. Front Chem. 2022;10:880193.
47. Bhatt DL, Szarek M, Steg PG, Cannon CP, Leiter LA, McGuire DK, et al. Sotagliflozin in patients with diabetes and recent worsening heart failure. N Engl J Med. 2021;384(2):117–28.
48. Anand P, Kunnumakkara AB, Newman RA, Aggarwal BB. Bioavailability of curcumin: problems and promises. Mol Pharm. 2007;4(6):807–18.
49. Lim H, Maltarollo VI, Trosset JY. Molecular modeling and molecular docking: challenges and applications to drug design. Front Mol Biosci. 2023;9:1032483.
50. Lotfi M, Naderi-Manesh H, Shams AR. Network pharmacology analysis of polyherbal anticancer combinations. Curr Pharm Biotechnol. 2022;23(14):1713–24.
51. Ross J, Belgodere B, Chenthamarakshan V, Padhi I, Mroueh Y, Das P. Large-scale chemical language representations capture molecular structure and properties. Nat Mach Intell. 2022;4(12):1256–64.